

The generalized work function algorithm is competitive for the generalized 2-server problem

René Sitters

VU University Amsterdam, The Netherlands.

Department of Operations Research r.a.sitters@vu.nl

November 1, 2011

Abstract

The generalized 2-server problem is an online optimization problem where a sequence of requests has to be served at minimal cost. Requests arrive one by one and need to be served instantly by at least one of two servers. We consider the general model where the cost function of the two servers may be different. Formally, each server moves in its own metric space and a request consists of one point in each metric space. It is served by moving one of the two servers to its request point. Requests have to be served without knowledge of the future requests. The objective is to minimize the total traveled distance. The special case where both servers move on the real line is known as the CNN problem. We show that the generalized work function algorithm WFA_λ is constant competitive for the generalized 2-server problem.

1 Introduction

The work function algorithm is a generic algorithm for online optimization problems. For many problems it gives the optimal competitive ratio or it is conjectured to be optimal. For example, it has the best known ratio of $2k - 1$ for the k -server problem, which is probably the most appealing and well-studied problem in online optimization, and the work function algorithm is conjectured to have an optimal ratio of k . There are many papers that deal with this classical work function algorithm. More powerful, but less known is the *generalized* work function algorithm, WFA_λ , which is the standard work function algorithm with an additional parameter λ . A result by Burley [8] shows that the general

algorithm can indeed be strictly more powerful than the standard work function algorithm.

The work function algorithm may be computationally expensive and pretty hard to analyze, but things can be much better for special cases. For example, the simple doubling algorithm for the cow path problem is in fact mimed by the generalized work function algorithm. Also, the complexity of the algorithm very much depends on the complexity of computing the offline solution. The work function algorithm for traversing layered graphs [8] can be implemented in linear time for instance. Further, the performance of the work function algorithm may be much better in practice than what is guaranteed in theory (See for example [5]). For some problems, the work function algorithm is optimal but there are more efficient alternatives. For example, it is k -competitive for weighted caching [4] but the elegant Double Coverage algorithm [10] has the same optimal ratio. However, the DC-algorithm is not extendable to arbitrary metric spaces. For some hard problems, the work function algorithm is basically the only algorithm known. Examples are the (deterministic) k -server problem and the generalized 2-server problem that we discuss in this paper.

In the *generalized 2-server problem* we are given a server, whom we will call the \mathbb{X} -server, moving in a symmetric metric space \mathbb{X} , and a server, the \mathbb{Y} -server, moving in a symmetric metric space \mathbb{Y} . A starting point $(\mathcal{O}^{\mathbb{X}}, \mathcal{O}^{\mathbb{Y}}) \in \mathbb{X} \times \mathbb{Y}$ is given and requests $(x, y) \in \mathbb{X} \times \mathbb{Y}$ are presented on-line one by one. Requests are served by moving one of the servers to the corresponding point in its metric space and the choice of which server to move is made without knowledge of the future requests. The objective is to minimize the sum of the distances traveled by the two servers. The special case $\mathbb{X} = \mathbb{Y} = \mathbb{R}$ is known as the CNN problem [24][25]. This problem can be seen as a single server moving in \mathbb{R}^2 and each request is a point in \mathbb{R}^2 which is served if the x - or y -coördinate of the server and request coincide.

Research on the CNN problem started more than ten years ago but despite the simplicity of the problem and its importance for the theory of online optimization, the problem is still not well-understood. The CNN problem first appeared in a paper of Koutsoupias and Taylor [24, 25]. They conjectured that the generalized work function algorithm WFA_λ has a constant competitive ratio¹ for any $\lambda \in (0, 1)$. They also conjectured that the generalized work function algorithm is competitive for the generalized 2-server problem. In this paper we settle both conjectures. The constant that follows from our proof is large and we do not present an upper bound on its value. Hence, the gap between known lower and upper bound remains large.

¹The λ in [25] corresponds with $1/\lambda$ in our notation.

We say that an algorithm ALG for an online minimization problem is *c-competitive* ($c \geq 1$) if there is a constant c_0 such that for every instance \mathcal{I} of the problem, the algorithms cost $\text{ALG}(\mathcal{I})$ and the optimal cost $\text{OPT}(\mathcal{I})$ satisfy

$$\text{ALG}(\mathcal{I}) \leq c \cdot \text{OPT}(\mathcal{I}) + c_0.$$

The competitive ratio of the algorithm is the infimum over c such that ALG is c -competitive. The competitive ratio of the online minimization problem is the infimum over c such that there is a c -competitive algorithm.

As the name suggests, the generalized 2-server problem originates from the *classical* 2-server problem in which $\mathbb{X} = \mathbb{Y}$ and $x = y$ for every request, i.e., each request is a point in the metric space and we have to decide which server to move to the requested point. The k -server problem (with $k \geq 2$ servers) is one of the most studied problems in online optimization. A review of the k -server problem by Koutsoupias appeared recently [21]. The k -server problem on a uniform metric space is the *paging problem*. In the *weighted k-server* problem a weight is assigned to each server (of the classical problem) and the total cost is the weighted sum of the distances. The weighted k -server problem is a special case of the generalized k -server problem.

All online optimization problems mentioned in this article belong to the class of *metrical task systems* (For a definition see Section 1.3 and [6]). Given multiple metrical task systems the *sum problem* [25] is again a metrical task system and is defined as follows: At each step we receive one request for each task system and we have to serve at least one of those requests. The CNN problem is the sum of two trivial problems: in both problems there is one server moving on the real line and each request consists of a single point. Koutsoupias and Taylor [25] emphasize the importance of the CNN problem: ‘*It is a very simple sum problem, which may act as a stepping stone towards building a robust (and less ad hoc) theory of online computation*’. Indeed, our techniques are useful for sum problems in general and we hope it leads to a better insight and hence further simplifications and generalizations of the theory of online computation.

1.1 Known competitive ratios

No *memoryless* (randomized) algorithm has a finite competitive ratio for the CNN problem [15],[25],[31] while a finite ratio is possible if we are allowed to store the entire given sequence (or at least the current work function) [28, 29]. The algorithm in both these papers is complex and the ratio very high but they apply to the generalized 2-server problem as well. See [9] for a review of [29]. For the classical k -server problem, the work function algorithm is $2k - 1$ competitive for any metric space [23] and it is conjectured to be even k -competitive [26,

23]. This famous *k-server conjecture* was posed more than two decades ago and is still open. It has been very influential on the development of the area online algorithms. The competitive ratio of the weighted *k*-server problem is much higher. Fiat and Ricklin [16] prove that for any metric space with at least $k + 1$ points there exists a set of weights such that the competitive ratio of any deterministic algorithm is at least $k^{\Omega(k)}$. Koutsoupias and Taylor [23] prove that any deterministic online algorithm for the weighted 2-server problem has competitive ratio at least $6 + \sqrt{17} > 10.12$ even if the underlying metric space is the line and [15] shows that any memoryless randomized algorithm has unbounded competitive ratio in this case. Consequently, both lower bounds apply to the CNN problem as well.

1.2 More special cases and variants

The *orthogonal* CNN problem [20] is the special case of the CNN problem in which each request either shares the x -coordinate or the y -coordinate with the previous request. Iwama and Yonezawa [20] give a 9-competitive algorithm and a lower bound of 3 is given in [1].

In the *continues* CNN problem [1] there is one request which follows a continuous path in \mathbb{R}^2 and the online server must serve it continuously by aligning either horizontally or vertically. It generalizes the orthogonal version in the sense that any c -competitive algorithm for the continues problem implies a c -competitive algorithm for the orthogonal problem. Augustine and Gravin [1] give a 6.46-competitive memoryless algorithm (improving the 9 from [20]).

The *axis-bound* CNN problem was introduced by Iwama and Yonezawa [18, 19] and is the special case in which the server can only move on the x - and y -axis. They give an upper bound of 9 and a lower bound of $4 + \sqrt{5}$. The lower bound was raised to 9 in [3]. In that paper they also give an alternative 9-competitive algorithm by formulating it as a *two point request problem* [8]. Finally, the *box bound* CNN problem [2] is the restriction in which the server can move only on the boundary of a rectangle and requests are inside the rectangle. The problem can be transformed into the 4-point request problem [2]. An upper bound of 88.71 for the latter problem follows from the paper by Burley [8].

For the weighted 2-server problem the only competitive algorithm follows from the one for the generalized 2-server problem. For the special case of a *uniform* metric space (where all distances are 1) Chrobak and Sgall [15] prove that the work function algorithm is 5-competitive and that no better ratio is possible. They also give a 5-competitive randomized, memoryless algorithm for uniform spaces, and a matching lower bound. Further, they consider a version of the problem in which a request specifies two points to be covered

by the servers, and the algorithm must decide which server to move to which point. For this version, they show a 9-competitive algorithm and prove that no better ratio is possible. Finally, Verhoeven [31] shows that no memoryless randomized algorithm can be competitive for the CNN problem under an even *weaker definition* of memoryless than used in [14] and [24].

1.3 Metrical task systems and metrical service systems

Borodin, Linial, and Saks [6] introduced the problem of *metrical task systems*, a generalization of all online problems discussed here. Such system is a pair $\mathcal{S} = (\mathbb{M}, \mathcal{T})$, where \mathbb{M} is a metric space and \mathcal{T} a set of tasks. Each task $\tau \in \mathcal{T}$ is defined by a function $\tau : \mathbb{M} \rightarrow \mathbb{R}^+$ which gives for each $s \in \mathbb{M}$ the cost of serving the task while being in s . In an online instance the tasks are given one by one and the objective is to minimize the total traveled distance (starting from given origin \mathcal{O}) plus the total service cost. The system is called *unrestricted* if \mathcal{T} consists of all non-negative real functions on \mathbb{M} . The authors of [6] show that the competitive ratio is exactly $2m - 1$ for the unrestricted metrical task system on any metric space on m points.

A restricted model is that of *metrical service systems*, introduced in [11],[12] and [27]. (In [27] it is called forcing task systems.) Such a system is a pair $\mathcal{S} = (\mathbb{M}, \mathcal{R})$, where \mathbb{M} is a metric space and \mathcal{R} a set of requests where each request $r \in \mathcal{R}$ is a subset of \mathbb{M} . The system is called *unrestricted* if \mathcal{R} consists of all subsets of \mathbb{M} . Metrical service systems correspond to metrical task systems for which $\tau : \mathbb{M} \rightarrow \{0, \infty\}$ for each task τ . Manasse et al. [27] give an optimal $m - 1$ -competitive algorithm for the unrestricted metrical service system on any metric space on m points.

The generalized 2-server problem is a metrical service system: There is one server moving in the product space $\mathbb{M} = \mathbb{X} \times \mathbb{Y}$ and any pair $(x, y) \in \mathbb{X} \times \mathbb{Y}$ defines a request $r(x, y) = \{\{x\} \times \mathbb{Y}\} \cup \{\mathbb{X} \times \{y\}\} \subset \mathbb{M}$. The distance between points (x_1, y_1) and (x_2, y_2) in $\mathbb{X} \times \mathbb{Y}$ is $d((x_1, y_1), (x_2, y_2)) = d^{\mathbb{X}}(x_1, x_2) + d^{\mathbb{Y}}(y_1, y_2)$, where $d^{\mathbb{X}}$ and $d^{\mathbb{Y}}$ are the distance functions of the metric spaces \mathbb{X} and \mathbb{Y} .

The work function algorithm is optimal for metrical task and metrical service systems in the sense that it is, respectively, $2m - 1$ and $m - 1$ competitive on any metric space of at most m points [13]. This is not of direct use for the CNN problem since the metric space, \mathbb{R}^2 , has an unbounded number of points.

1.4 The work function algorithm: WFA_λ

The work function algorithm appeared for the first time in [11] but was discovered independently by others (see [23]). We use it here only for metrical service

systems but it works the same for metrical task systems.

Definition 1 *Given a metrical service system $\mathcal{S} = (\mathbb{M}, \mathcal{R})$ and origin $\mathcal{O} \in \mathbb{M}$, and given a request sequence σ the work function $W_\sigma : \mathbb{M} \rightarrow \mathbb{R}^+$ is defined as follows. For any point $s \in \mathbb{M}$, $W_\sigma(s)$ is the length of the shortest path that starts in \mathcal{O} , ends in s and serves σ .*

We assume here that the work function is well-defined, which might not be true if the metric space is infinite. Thus, we assume that for any $\sigma = r_1, \dots, r_n$ and any point $s \in \mathbb{M}$ there are points $s_i \in r_i$ ($i = 1, \dots, n$) such that $d(\mathcal{O}, s_1) + d(s_1, s_2) + \dots + d(s_{n-1}, s_n) + d(s_n, s) \leq d(\mathcal{O}, t_1) + d(t_1, t_2) + \dots + d(t_{n-1}, t_n) + d(t_n, s)$ for any set of points $t_i \in r_i$ ($i = 1, \dots, n$). Clearly, the work function is well-defined for the generalized 2-server problem since we may assume that for each s_i , both *coördinates* are from requests given sofar, thus the number of interesting paths is finite. See [12] for a sufficient condition for the work function to be well-defined.

For a work function W_σ we say that point s is *dominated* by point t if $W_\sigma(s) = W_\sigma(t) + d(s, t)$. We define the *support* of W_σ as $\text{supp}(W_\sigma) = \{s \in \mathbb{M} : s \text{ is not dominated by any other point}\}$. If $W_{\sigma,r}$ is a well-defined work function then $\text{supp}(W_{\sigma,r}) \subseteq r$ since for any point $s \notin r$ there exists a point $t \in r$ such that $W_{\sigma,r}(s) = W_{\sigma,r}(t) + d(t, s)$. For more properties and a deeper analysis of the work function (algorithm) see for example [7],[8], and [21].

The generalized work function algorithm is a work function-based algorithm parameterized by some constant $\lambda \in (0, 1]$. We denote it by WFA_λ .

Definition 2 *For any request sequence σ and any new request r , the generalized work function algorithm WFA_λ moves the server from the position s it had after serving σ to any point*

$$s' \in \text{Argmin}_{t \in \mathbb{M}} \{W_{\sigma,r}(t) + \lambda d(s, t)\}. \quad (1)$$

This minimum may not be well-defined if the request r contains infinitely many points of the metric space. This is no problem for the generalized 2-server problem since the minimum is attained for some t with both *coördinates* of the given requests [28]. From (1) we see that

$$W_{\sigma,r}(s') + \lambda d(s, s') \leq W_{\sigma,r}(t) + \lambda d(s, t) \quad \text{for any point } t \in \mathbb{M}.$$

Using the triangle inequality we get that for any $t \in \mathbb{M}$

$$W_{\sigma,r}(s') \leq W_{\sigma,r}(t) + \lambda(d(s, t) - d(s, s')) \leq W_{\sigma,r}(t) + \lambda d(s', t). \quad (2)$$

If $\lambda < 1$ then (2) implies that s' is not dominated by any other point, whence $s' \in \text{supp}(W_{\sigma,r}) \subseteq r$. We see that if the moves of WFA_λ are well-defined then

the choice of $\lambda < 1$ ensures that the point s' always serves the last request and we may *replace* $t \in \mathbb{M}$ by $t \in r$ in Definition 2.

For $\lambda = 0$, the generalized work function algorithm corresponds to the algorithm that always moves to the endpoint of an optimal solution, and for $\lambda = \infty$ it corresponds to the greedy algorithm (if we take $t \in r$ in stead of $t \in \mathbb{M}$ in (1)). The standard work function algorithm has $\lambda = 1$ and was first used in [11] and has been studied extensively. The general form was defined in [11] as well but was used only shortly after in [12] where it is called the λ -Cheap-and-Lazy strategy. They show that WFA_λ with $\lambda = 1/3$ is optimal for the 2-point request problem. Burley generalized this and showed that WFA_λ is $O(k2^k)$ -competitive for the k -point request problem (where λ depends on k).

Usually, λ is placed before $W_{\sigma,r}$ in (1) in stead of before $d(s,t)$, as we do here. Also, sometimes α is used in stead of λ . For example, Burley [8] uses $\alpha > 1$ and the following definition of the work function algorithm:

$$s' \in \text{Argmin}_{t \in \mathbb{M}} \{\alpha W_{\sigma,r}(t) + d(s,t)\}.$$

Replacing α by $1/\lambda$ matches our definition. Our choice was purely for an aesthetical reason: Now the term λ appears much more often in the paper than $1/\lambda$. Having said that, $\lambda < 1$ indeed seems the better way, especially considering the definitions of extended cost and slack function (See Section 2).

1.5 Paper outline and proof sketch

The major part of this paper is devoted to the CNN problem (Theorem 1). The generalization to arbitrary metric spaces (Theorem 2) is more complex and we do this in a separate section. The proof of Theorem 1 is based on no less than 21 lemmas. To obtain a better insight in the relation between lemmas we mention after each lemma where it is used. The proof of Theorem 2 uses exactly the same lemmas, only some constants are different. We indicate how to adjust the proofs. Before giving a sketch of the proof we give a brief outline of the paper.

In Section 2 we list some properties of the work function algorithm. These hold for any metrical service system and can be found in several other papers, e.g. [8, 13, 23]. Further, we introduce the closely related slack function and list some of its properties. In Section 3 we present our potential function for the CNN problem together with some of its properties. Although the potential function is defined for the CNN problem, the theory in Sections 3.1 and 3.2 applies to any metrical service system on \mathbb{R}^2 . In Section 3.3 we state some properties of the CNN problem which do not depend on the potential and in Section 3.4 we put everything together and apply the potential to the CNN problem. In Section 4 we show how to modify the proof for general metric spaces. There are several

reasons for giving a separate CNN proof. First, the reader has the option of just reading the CNN proof and skip the more difficult general proof. Nevertheless, we believe that the generalization is relatively easy to digest once the reader has worked through the CNN proof and it may be even easier this way than when we would present only the general proof. One reason is that the CNN problem can be seen as moving points in the Euclidean plane which makes the proof easier to visualize than the proof for the general case.

Our potential function has a long description and may seem unintuitive at first. It is a linear combination of two functions: \mathcal{F} and \mathcal{G} . Function \mathcal{F} is a special case of the potential function that was used in [28] to give the first constant competitive algorithm for the generalized 2-server problem. When we use only \mathcal{F} as our potential function and follow the line of proof that we use here, then the analysis fails. Taking \mathcal{G} as potential function doesn't work either. However, the two functions are in a way complementary and if we take a linear combination of the two functions then the proof goes through.

Next, we give a short technical sketch of the proof, which applies to both the CNN problem and the general problem. This part may be unclear for some readers but may be very helpful for readers that are familiar with analysis of the work function algorithm. The potential function has the following form.

$$\Phi_\sigma = (1 - \gamma) \min_{s_1, s_2, s_3 \in \mathbb{M}} \mathcal{F}_\sigma(s_1, s_2, s_3) + \gamma \min_{s_1, s_2, s_3 \in \mathbb{M}} \mathcal{G}_\sigma(s_1, s_2, s_3),$$

where $\mathbb{M} = \mathbb{X} \times \mathbb{Y}$. The initial value is zero and in general it is upper bounded by the optimal value of the instance. We consider two subsequent requests r' and r'' and show that the increase $\Phi'' - \Phi'$ is at least some constant times the so called extended cost for r'' . Proof of competitiveness then follows directly.

Let us give some more details. Let σ' be request sequence that ends with r' . It is followed by r'' and we denote $\sigma'' = \sigma', r''$. Let s_1, s_2, s_3 be a minimizer of $\mathcal{F}_{\sigma''}$. By construction of \mathcal{F} , all three points will serve the last request r'' . (The same holds for $\mathcal{G}_{\sigma''}$.) We distinguish between the case $|\{s_1, s_2, s_3\}| \leq 2$ and the case $|\{s_1, s_2, s_3\}| = 3$. In the following, $c_1, c_2, c_3, c_4 > 0$ are specific constants depending on λ . For the first case we show that

$$\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} \geq c_1 \nabla_{r''} \quad \text{and} \quad \min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} \geq 0,$$

where $\nabla_{r''}$ is the extended cost for r'' w.r.t. σ' . Hence, the increase for Φ is at least $(1 - \gamma)c_1 \nabla_{r''}$, which is constant times the extended cost and the proof follows. For the second case we show that

$$\begin{aligned} \min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} &\geq c_2 \min\{\partial x, \partial y\} \quad \text{and} \\ \min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} &\geq c_3 \nabla_{r''} - c_4 \min\{\partial x, \partial y\}, \end{aligned}$$

where $\partial x = d^{\mathbb{X}}(x', x'')$ and $\partial y = d^{\mathbb{Y}}(y', y'')$. Hence, the increase in Φ in this case is at least

$$\gamma c_3 \nabla_{r''} + ((1 - \gamma)c_2 - \gamma c_4) \min\{\partial x, \partial y\},$$

which is at least $\gamma c_3 \nabla_{r''}$ for γ small enough.

In Section 5.1 we give a sketch of a possible extension to higher dimensions.

2 Preliminaries

For the analysis of the generalized work function algorithm we make extensively use of two concepts: extended cost and slack. The first is an amortized cost of the (general) work function algorithm. It was introduced together with the work function algorithm in [11] (where it is called pseudo cost) and has been used in every analysis of the work function algorithm. The slack function was defined by Burley [8] and was also used in [28]. Its definition comes naturally with that of extended cost and its use enhances the analysis. This section applies to any metrical service system.

2.1 The extended cost

Definition 3 For request sequence σ and request r the extended cost for r is

$$\nabla_r(W_\sigma) = \max_{s \in \mathbb{M}} \min_{t \in r} [W_\sigma(t) + \lambda d(s, t) - W_\sigma(s)].$$

For $\rho = r_1 r_2 \cdots r_n$ we define the total extended cost as $\nabla_\rho = \sum_{i=1}^n \nabla_{r_i}(W_{r_1 \cdots r_{i-1}})$.

The definition of extended cost matches that in [28] and matches the commonly used extended cost in case $\lambda = 1$. It also matches the definition by Burley [8], although the notation is quite different. The intuition behind extended cost becomes clear from the following lemma and its proof.

In this paper, ρ always refers to the entire given sequence, i.e., no requests are given after ρ . We mainly use σ otherwise.

Lemma 1 If $\nabla_\rho \leq c \text{OPT}_\rho$ for some constant c and any request sequence ρ , then WFA_λ is $(c - 1)/\lambda$ -competitive. (Used in proof of Theorem 1)

Proof: Assume the online server is in point s' after it served the initial sequence σ and moves to t' to serve a new request r . Since we maximize over $s \in \mathbb{M}$ in Definition 3 we have

$$\nabla_r(W_\sigma) \geq \min_{t \in r} [W_\sigma(t) + \lambda d(s', t) - W_\sigma(s')].$$

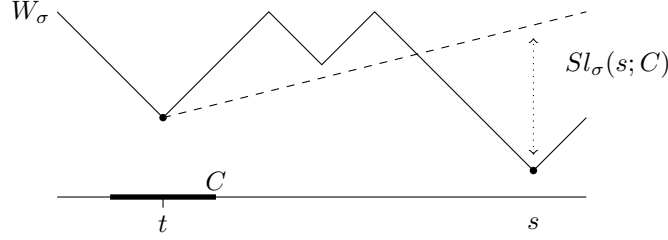


Figure 1: Slack: A possible work function over a line segment. The slack of s to C w.r.t. σ is attained in $t \in C$.

By definition of WFA_λ the minimum in the right side is attained for $t = t'$. Therefore,

$$\nabla_r(W_\sigma) \geq W_\sigma(t') + \lambda d(s', t') - W_\sigma(s').$$

Rewriting we get

$$d(s', t') \leq \frac{1}{\lambda} (\nabla_r(W_\sigma) + W_\sigma(s') - W_\sigma(t')).$$

This gives an upper bound for the cost of WFA_λ for serving some single request r . Let q be the point where the algorithm ends after serving ρ . Summing-up over all requests in ρ we get that the total cost for WFA_λ is at most

$$\frac{1}{\lambda} (\nabla_\rho + W_\epsilon(\mathcal{O}) - W_\rho(q)) \leq \frac{1}{\lambda} (\nabla_\rho - \text{OPT}_\rho) \leq \frac{1}{\lambda} (c - 1) \text{OPT}_\rho.$$

□

2.2 The slack function

We use the concept of the *slack* of a point relative to another point. Intuitively, the slack of a point s with respect to a point t is the amount that the work function value in s can increase before the generalized work function algorithm moves from s to t . More precisely, the generalized work function algorithm, being in point s after serving sequence σ , moves away from s after a new request r is given if there is a point t such that $W_{\sigma,r}(t) + \lambda d(s, t) \leq W_{\sigma,r}(s)$. The slack is the difference between the left and right side of this inequality. More generally, we define the slack of a point to a subset of \mathbb{M} . See Figure 1.

Definition 4 *Given a request sequence σ we define the slack of a point $s \in \mathbb{M}$ to a (possibly infinite) set of points $C \subseteq \mathbb{M}$ and with respect to σ as*

$$Sl_\sigma(s; C) = \min_{t \in C} \{W_\sigma(t) + \lambda d(t, s)\} - W_\sigma(s).$$

If C contains only one point t then we simply write $Sl_\sigma(s; t)$ instead of $Sl_\sigma(s; \{t\})$.

Using the slack function makes the proof shorter and more intuitive. For example, we can rewrite the extended cost, $\nabla_r(W_\sigma)$, for request sequence σ and new request r in terms of the slack function.

$$\begin{aligned}\nabla_r(W_\sigma) &= \max_{s \in \mathbb{M}} \{ \min_{t \in r} \{ W_\sigma(t) + \lambda d(s, t) \} - W_\sigma(s) \} \\ &= \max_{s \in \mathbb{M}} Sl_\sigma(s; r).\end{aligned}\tag{3}$$

In the remainder of this section we list some properties of the slack function. The first follows directly from its definition and from the work function being Lipschitz continuous with constant 1. For any $s, t \in \mathbb{M}$

$$Sl_\sigma(s; t) \leq (1 + \lambda)d(s, t).\tag{4}$$

The next lemma also follows directly from the definition.

Lemma 2 *If $C_1 \subseteq C_2 \subseteq \mathbb{M}$, then for any $s \in \mathbb{M}$ we have $Sl_\sigma(s; C_1) \geq Sl_\sigma(s; C_2)$. (Used in proof of many lemmas.)*

Lemma 2 is mostly used in the form: $t \in C \subseteq \mathbb{M}$ implies $Sl_\sigma(s; t) \geq Sl_\sigma(s; C)$.

Lemma 3 *For any set of points $C \subset \mathbb{M}$ there is a points $s \in C$ such that $Sl_\sigma(s; C) = 0$. (Used in proof of Lemma 9.)*

Proof: Let $s \in \text{Argmin}\{W_\sigma(t) \mid t \in C\}$. Then, for any $t \in C$:

$$Sl_\sigma(s; t) = W_\sigma(t) + \lambda d(t, s) - W_\sigma(s) \geq 0.$$

Clearly, $Sl_\sigma(s; s) = 0$. Hence, $Sl_\sigma(s; C) = \min_{t \in C} Sl_\sigma(s; t) = 0$. \square

The next lemma shows a transitivity property of slack.

Lemma 4 *Let $s_1, s_2, s_3 \in \mathbb{M}$ such $d(s_1, s_2) + d(s_2, s_3) = d(s_1, s_3)$. Then $Sl_\sigma(s_3; s_1) = Sl_\sigma(s_3; s_2) + Sl_\sigma(s_2; s_1)$. (Used in Lemma 11.)*

Proof:

$$\begin{aligned}Sl_\sigma(s_3; s_1) &= W_\sigma(s_1) + \lambda d(s_1, s_3) - W_\sigma(s_3) \\ &= W_\sigma(s_1) + \lambda(d(s_1, s_2) + d(s_2, s_3)) - W_\sigma(s_3) \\ &= W_\sigma(s_1) + \lambda d(s_1, s_2) - W_\sigma(s_2) + W_\sigma(s_2) + \lambda d(s_2, s_3) - W_\sigma(s_3) \\ &= Sl_\sigma(s_2; s_1) + Sl_\sigma(s_3; s_2).\end{aligned}$$

\square

The next lemma generalizes Lemma 2.

Lemma 5 Let $C_1, C_2 \subseteq \mathbb{M}$ and $\delta \in \mathbb{R}^+$. If for every point $u_1 \in C_1$ there is a point $u_2 \in C_2$ with $d(u_1, u_2) \leq \delta$, then for every $s \in \mathbb{M}$

$$Sl_\sigma(s; C_1) \geq Sl_\sigma(s; C_2) - (1 + \lambda)\delta.$$

(Used in proof of Lemma 7.)

Proof: Let $u_1 \in C_1$ be such that $Sl_\sigma(s; C_1) = Sl_\sigma(s; u_1)$. There is a point $u_2 \in C_2$ such that $d(u_1, u_2) \leq \delta$.

$$\begin{aligned} & Sl_\sigma(s; C_1) - Sl_\sigma(s; C_2) \\ &= Sl_\sigma(s; u_1) - Sl_\sigma(s; C_2) \\ &\geq Sl_\sigma(s; u_1) - Sl_\sigma(s; u_2) \\ &= W_\sigma(u_1) + \lambda d(u_1, s) - W_\sigma(s) - (W_\sigma(u_2) + \lambda d(u_2, s) - W_\sigma(s)) \\ &= W_\sigma(u_1) - W_\sigma(u_2) + \lambda(d(u_1, s) - d(u_2, s)) \\ &\geq W_\sigma(u_1) - W_\sigma(u_2) - \lambda d(u_1, u_2) \\ &\geq -d(u_1, u_2) - \lambda d(u_1, u_2) \\ &= -(1 + \lambda)\delta. \end{aligned}$$

□

Lemma 6 Let $s, t \in \mathbb{M}$ and $C \subset \mathbb{M}$. Then,

- (a) $Sl_\sigma(t; C) \geq Sl_\sigma(s; C) - (1 + \lambda)d(s, t)$, and
- (b) $Sl_\sigma(t; C) \geq Sl_\sigma(s; C) + (1 - \lambda)d(s, t)$, if t dominate s w.r.t. σ .

(Used in proof of Lemma 7, 15 and 19.)

Proof: Let $u \in C$ be such that $Sl_\sigma(t; C) = Sl_\sigma(t; u)$. Then,

$$\begin{aligned} Sl_\sigma(t; C) &= Sl_\sigma(t; u) \\ &= Sl_\sigma(s; u) - \lambda d(u, s) + \lambda d(u, t) + W_\sigma(s) - W_\sigma(t) \\ &\geq Sl_\sigma(s; u) - \lambda d(s, t) + W_\sigma(s) - W_\sigma(t) \\ &\geq Sl_\sigma(s; C) - \lambda d(s, t) + W_\sigma(s) - W_\sigma(t). \end{aligned}$$

The first inequality is given by the triangle inequality and the second by Lemma 2. In general, $W_\sigma(s) - W_\sigma(t) \geq -d(s, t)$. If t dominate s then we have the stronger bound $W_\sigma(s) - W_\sigma(t) = d(s, t)$. □

3 The CNN problem

A simple example shows that the standard work function algorithm has unbounded competitive ratio for the CNN problem.

Example 1 *The standard work function algorithm WFA_1 is not competitive for the CNN problem. Take $(0,0)$ as the origin and consider the request sequence $(1,2), (2,2), (3,2), \dots, (m,2)$ for arbitrary m . The work function algorithm follows the path $(0,0), (1,0), (2,0), \dots, (m,0)$. (There are no draws.) The optimal path is $(0,0), (0,2)$. The competitive ratio for this instance is $m/2$.*

Theorem 1 *The generalized work function algorithm WFA_λ is constant competitive for the CNN problem for any constant λ with $0 < \lambda < 1$.*

All the lemmas of the previous section apply to metrical service systems in general. In this section we restrict to the CNN problem. It is convenient to insist on writing \mathbb{M} for the metric space although we now have $\mathbb{M} = \mathbb{R}^2$. We make a subtle distinction between the request point $(x', y') \in \mathbb{R}^2$ and the corresponding request $r(x', y') = \{(x, y) \in \mathbb{R}^2 \mid x = x' \text{ or } y = y'\}$.

3.1 The potential function

Our potential function is defined for any metrical service system on \mathbb{R}^2 but we only use it for the CNN problem.

One of the ingredients is the set $\text{BOX}(s_1, s_2)$ (see Figure 2) defined as follows. Given points $x_1, x_2 \in \mathbb{R}$ we denote by $[x_1, x_2]$ the interval between x_1 and x_2 (we allow $x_2 < x_1$, i.e., $[x_2, x_1] = [x_1, x_2]$). Note that here we use the restriction to the real line since this is not well-defined for a general metric space. Given points $s_1 = (x_1, y_1) \in \mathbb{M}$ and $s_2 = (x_2, y_2) \in \mathbb{M}$ we denote the set of points in the rectangle spanned by these points by

$$\text{BOX}(s_1, s_2) = \{(x, y) \in \mathbb{M} \mid x \in [x_1, x_2] \text{ and } y \in [y_1, y_2]\}.$$

Let $0 < \alpha < 1/2$ and $0 < \gamma < 1$. We define the functions $\mathcal{F}_\sigma : \mathbb{M}^3 \rightarrow \mathbb{R}$ and $\mathcal{G}_\sigma : \mathbb{M}^3 \rightarrow \mathbb{R}$ as

$$\begin{aligned} \mathcal{F}_\sigma(s_1, s_2, s_3) &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \alpha Sl_\sigma(s_3; \{s_1, s_2\}) \\ \mathcal{G}_\sigma(s_1, s_2, s_3) &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \alpha Sl_\sigma(s_3; \text{BOX}(s_1, s_2)). \end{aligned}$$

The two functions only differ in the last term. The potential function Φ_σ is

$$\Phi_\sigma = (1 - \gamma) \min_{s_1, s_2, s_3 \in \mathbb{M}} \mathcal{F}_\sigma(s_1, s_2, s_3) + \gamma \min_{s_1, s_2, s_3 \in \mathbb{M}} \mathcal{G}_\sigma(s_1, s_2, s_3).$$

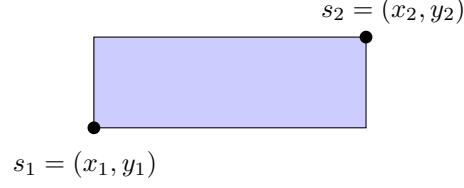


Figure 2: The shaded area is $\text{BOX}(s_1, s_2)$, used in the potential function.

The numbers α and γ will depend only on λ and we fix their precise values later. It is good to mention here that the proof works for any small enough values of α and γ . More precisely, the proof works if we pick any α with $0 < \alpha \leq \alpha_0$ for some α_0 depending on λ and then pick any γ with $0 < \gamma \leq \gamma_0$ for some γ_0 depending on λ and α .

Comprehensive notation

To simplify the analysis we define one more function $\mathcal{H}_\sigma : \mathbb{M}^2 \rightarrow \mathbb{R}$. It corresponds with the first two terms of \mathcal{F}_σ and \mathcal{G}_σ .

$$\mathcal{H}_\sigma(s_1, s_2) = W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) = \frac{1}{2}W_\sigma(s_1) + \frac{1}{2}W_\sigma(s_2) - \frac{\lambda}{2}d(s_1, s_2).$$

We can rewrite \mathcal{F}_σ and \mathcal{G}_σ as

$$\mathcal{F}_\sigma(s_1, s_2, s_3) = \mathcal{H}_\sigma(s_1, s_2) - \alpha Sl_\sigma(s_3; \{s_1, s_2\}),$$

$$\mathcal{G}_\sigma(s_1, s_2, s_3) = \mathcal{H}_\sigma(s_1, s_2) - \alpha Sl_\sigma(s_3; \text{BOX}(s_1, s_2)).$$

For a request sequence σ we denote $\min_{s_1, s_2, s_3 \in \mathbb{M}} \mathcal{F}_\sigma(s_1, s_2, s_3)$ simply by $\min \mathcal{F}_\sigma$ and make a similar simplification of notation for \mathcal{G} and \mathcal{H} . A shorter notation for the potential function becomes

$$\Phi_\sigma = (1 - \gamma) \min \mathcal{F}_\sigma + \gamma \min \mathcal{G}_\sigma.$$

Note that \mathcal{H}_σ is symmetric in s_1 and s_2 and, consequently, also \mathcal{F}_σ and \mathcal{G}_σ are symmetric in s_1 and s_2 . This property is not essential but enhances the argumentation at some points.

3.2 Properties of the potential function

In this section we list some properties of the potential function Φ_σ which hold for any metrical service system on $\mathbb{M} = \mathbb{R}^2$ and arbitrary corresponding request sequence σ . In section 3.3 we restrict the analysis to the CNN problem.

The functions \mathcal{F}_σ and \mathcal{G}_σ are constructed such that in the minimum all three points s_1, s_2, s_3 are on the last request, at least if α is small enough. This is stated in Lemma 8. The next lemma is preliminary for this lemma and several others.

Lemma 7 *Let $t \in \mathbb{M}$ dominate $s \in \mathbb{M}$ (w.r.t. σ) and let $\delta = d(s, t)$. Then, for any $s_1, s_2, s_3 \in \mathbb{M}$*

- (a) $\mathcal{F}_\sigma(s_1, s_2, s) - \mathcal{F}_\sigma(s_1, s_2, t) \geq \delta \cdot \alpha(1 - \lambda)$
- (b) $\mathcal{F}_\sigma(s_1, s, s_3) - \mathcal{F}_\sigma(s_1, t, s_3) \geq \delta \cdot (\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda))$
- (c) $\mathcal{F}_\sigma(s, s_2, s_3) - \mathcal{F}_\sigma(t, s_2, s_3) \geq \delta \cdot (\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda))$
- (d) $\mathcal{G}_\sigma(s_1, s_2, s) - \mathcal{G}_\sigma(s_1, s_2, t) \geq \delta \cdot \alpha(1 - \lambda)$
- (e) $\mathcal{G}_\sigma(s_1, s, s_3) - \mathcal{G}_\sigma(s_1, t, s_3) \geq \delta \cdot (\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda))$
- (f) $\mathcal{G}_\sigma(s, s_2, s_3) - \mathcal{G}_\sigma(t, s_2, s_3) \geq \delta \cdot (\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda))$

(Used in proof of Lemma 8, 12, 17, 18, and 19.)

Proof: Statements (a) and (d) follow directly from Lemma 6(b). By symmetry of \mathcal{F} and \mathcal{G} in their first two arguments it only remains to prove statements (b) and (e). We start with (b).

$$\begin{aligned} & \mathcal{F}_\sigma(s_1, s, s_3) - \mathcal{F}_\sigma(s_1, t, s_3) \\ &= \frac{1}{2}(Sl_\sigma(t; s_1) - Sl_\sigma(s; s_1)) + \alpha(Sl_\sigma(s_3; \{s_1, t\}) - Sl_\sigma(s_3; \{s_1, s\})). \end{aligned}$$

For the first part we use Lemma 6(b):

$$Sl_\sigma(t; s_1) - Sl_\sigma(s; s_1) \geq (1 - \lambda)\delta.$$

For the second part we apply Lemma 5 with $C_1 = \{s_1, t\}$ and $C_2 = \{s_1, s\}$. The condition of Lemma 5 is satisfied for $\delta = d(s, t)$. We have

$$Sl_\sigma(s_3; \{s_1, t\}) - Sl_\sigma(s_3; \{s_1, s\}) \geq -(1 + \lambda)\delta.$$

Combining those we get

$$\mathcal{F}_\sigma(s_1, s, s_3) - \mathcal{F}_\sigma(s_1, t, s_3) \geq \frac{1}{2}(1 - \lambda)\delta + \alpha(-(1 + \lambda)\delta).$$

The proof of (d) is similar. We apply Lemma 5 with $C_1 = \text{Box}(s_1, t)$ and $C_2 = \text{Box}(s_1, s)$. The condition of Lemma 5 is satisfied for $\delta = d(s, t)$.

$$\begin{aligned} & \mathcal{G}_\sigma(s_1, s, s_3) - \mathcal{G}_\sigma(s_1, t, s_3) \\ &= \frac{1}{2}(Sl_\sigma(t; s_1) - Sl_\sigma(s; s_1)) + \alpha(Sl_\sigma(s_3; \text{Box}(s_1, t)) - Sl_\sigma(s_3; \text{Box}(s_1, s))) \\ &\geq \frac{1}{2}(1 - \lambda)\delta - \alpha(1 + \lambda)\delta. \end{aligned}$$

□

Note that all the right hand sides in Lemma 7 are strictly positive if $0 < \alpha < (1 - \lambda)/(2(1 + \lambda))$. We assume this from now on.

Lemma 8 *If \mathcal{F}_σ or \mathcal{G}_σ is minimized in (s_1, s_2, s_3) , then $s_1, s_2, s_3 \in r$. (Used in proof of Lemma 18 and 19.)*

Proof: Any point is dominated by a point in the last request. Take any triple of points in \mathbb{M} . If at least one of the points is not in r , then Lemma 7 tells us that we can replace it by a point of the last request, r , such that the values of \mathcal{F}_σ and \mathcal{G}_σ become strictly smaller. \square

Lemma 9 $\min \mathcal{F}_\sigma \leq \min \mathcal{G}_\sigma \leq \min \mathcal{H}_\sigma$. (Used in proof of Lemma 20.)

Proof: For any $s_1, s_2 \in \mathbb{M}$ there is a point s_3 such that $Sl_\sigma(s_3; \text{Box}(s_1, s_2)) = 0$. (See Lemma 3.) Hence, $\min \mathcal{G}_\sigma \leq \min \mathcal{H}_\sigma$.

For any $s_1, s_2, s_3 \in \mathbb{M}$ we have $Sl_\sigma(s_3; \{s_1, s_2\}) \geq Sl_\sigma(s_3; \text{Box}(s_1, s_2))$ since $\{s_1, s_2\} \subseteq \text{Box}(s_1, s_2)$. (See Lemma 2.) Therefore, $\min \mathcal{F}_\sigma \leq \min \mathcal{G}_\sigma$. \square

The two inequalities of Lemma 9 are only strict if the three points for which the minimum of \mathcal{F}_σ or \mathcal{G}_σ is attained are in a way different enough. For example, the next lemma implies that if the minimum of \mathcal{F}_σ is attained for (s_1, s_2, s_3) but they are not all different, then both inequalities are equalities. For \mathcal{G}_σ a stronger property holds. If s_1, s_2, s_3 are all on a line then the second inequality is an equality.

Lemma 10 *If $\{s_1, s_2, s_3\}$ has cardinality at most 2, then $\mathcal{H}_\sigma(u_1, u_2) \leq \mathcal{F}_\sigma(s_1, s_2, s_3)$ for some $u_1, u_2 \in \{s_1, s_2, s_3\}$. (Used in proof of Lemma 18 and 20.)*

Proof: If $Sl_\sigma(s_3; \{s_1, s_2\}) \leq 0$, then $\mathcal{H}_\sigma(s_1, s_2) \leq \mathcal{F}_\sigma(s_1, s_2, s_3)$. So assume opposite:

$$Sl_\sigma(s_3; \{s_1, s_2\}) > 0. \quad (5)$$

We cannot have $s_1 = s_3$ or $s_2 = s_3$, since this contradicts (5). Hence, we must have $s_1 = s_2$, which implies $Sl_\sigma(s_2; s_1) = 0$.

$$\begin{aligned} \mathcal{F}_\sigma(s_1, s_2, s_3) &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \alpha Sl_\sigma(s_3; \{s_1, s_2\}) \\ &= W_\sigma(s_1) - \alpha Sl_\sigma(s_3; \{s_1, s_2\}) \\ &> W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_3; \{s_1, s_2\}) \\ &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_3; s_1) \\ &= \mathcal{H}_\sigma(s_1, s_3). \end{aligned}$$

For the inequality we used (5) and $\alpha < 1/2$. \square

Lemma 10 applies also to \mathcal{G}_σ instead of \mathcal{F}_σ but we shall not use this. In addition, \mathcal{G}_σ has the following property.

Lemma 11 *If $s_1, s_2, s_3 \in \mathbb{M}$ have the same x - or y -coördinate. Then $\mathcal{H}_\sigma(u_1, u_2) \leq \mathcal{G}_\sigma(s_1, s_2, s_3)$ for some $u_1, u_2 \in \{s_1, s_2, s_3\}$. (Used in proof of Lemma 19.)*

Proof: We shall prove something stronger than we need as this hardly changes the proof: If one of the three points is contained in $\text{BOX}(\cdot, \cdot)$ defined by the other two points. Then $\mathcal{H}_\sigma(u_1, u_2) \leq \mathcal{G}_\sigma(s_1, s_2, s_3)$ for some $u_1, u_2 \in \{s_1, s_2, s_3\}$. The lemma is a special case of this.

The proof is similar to that of Lemma 10. If $Sl_\sigma(s_3; \text{BOX}(s_1, s_2)) \leq 0$, then $\mathcal{H}_\sigma(s_1, s_2) \leq \mathcal{G}_\sigma(s_1, s_2, s_3)$ and we are done. So assume the opposite:

$$Sl_\sigma(s_3; \text{BOX}(s_1, s_2)) > 0. \quad (6)$$

We cannot have $s_3 \in \text{BOX}(s_1, s_2)$ since that contradicts 6. Hence, either $s_1 \in \text{BOX}(s_2, s_3)$ or $s_2 \in \text{BOX}(s_1, s_3)$. By symmetry of \mathcal{H} and \mathcal{G} in their first two arguments we may assume the latter is true. Hence, $d(s_1, s_2) + d(s_2, s_3) = d(s_1, s_3)$. By Lemma 4

$$Sl_\sigma(s_3; s_1) = Sl_\sigma(s_3; s_2) + Sl_\sigma(s_2; s_1). \quad (7)$$

Now, we use $Sl_\sigma(s_3; s_2) \geq Sl_\sigma(s_3; \text{BOX}(s_1, s_2))$ (by Lemma 2), $\alpha < 1/2$, and (6).

$$\begin{aligned} \mathcal{G}_\sigma(s_1, s_2, s_3) &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \alpha Sl_\sigma(s_3; \text{BOX}(s_1, s_2)) \\ &\geq W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \alpha Sl_\sigma(s_3; s_2) \\ &> W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_2; s_1) - \frac{1}{2}Sl_\sigma(s_3; s_2) \\ &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma(s_3; s_1) \\ &= \mathcal{H}_\sigma(s_1, s_3). \end{aligned}$$

□

Initially, the potential function is zero and in general it is upper bounded by the optimal value of the given sequence. This is stated in the next two lemmas. Let ϵ be the empty request sequence.

Lemma 12 $\Phi_\epsilon = 0$. (Used in proof of Theorem 1.)

Proof: Any point s is dominated by the origin \mathcal{O} , w.r.t. the empty sequence. By Lemma 7 we see that $\min \mathcal{F}_\epsilon = \mathcal{F}_\epsilon(\mathcal{O}, \mathcal{O}, \mathcal{O}) = 0$ and $\min \mathcal{G}_\epsilon = \mathcal{G}_\epsilon(\mathcal{O}, \mathcal{O}, \mathcal{O}) = 0$. □

Lemma 13 $\Phi_\rho \leq \text{OPT}_\rho$, for any sequence ρ . (Used in proof of Theorem 1.)

Proof: Let q be the endpoint of an optimal solution for ρ . Then $W_\rho(q) = \text{OPT}_\rho$ and $\mathcal{F}_\rho(q, q, q) = \mathcal{G}_\rho(q, q, q) = W_\rho(q)$. Hence, $\min \mathcal{F}_\rho \leq W_\rho(q) = \text{OPT}_\rho$ and $\min \mathcal{G}_\rho \leq W_\rho(q) = \text{OPT}_\rho$.

$$\Phi_\rho = (1 - \gamma) \min \mathcal{F}_\rho + \gamma \min \mathcal{G}_\rho \leq (1 - \gamma) \text{OPT}_\rho + \gamma \text{OPT}_\rho = \text{OPT}_\rho.$$

□

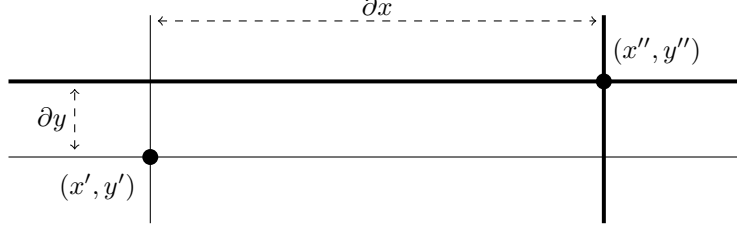


Figure 3: Two subsequent requests $r' = r(x', y')$ and $r'' = r(x'', y'')$. We assume $\partial x \geq \partial y$.

3.3 Properties of the CNN work function

Each metrical service system has its own specific properties of its work function. For example, Koutsoupias and Papadimitriou show a quasi-convexity property of the work function for the k -server problem [23]. A good understanding of the CNN work function is lacking but the two simple properties we show in this section are enough to prove constant competitiveness. Let σ' be an arbitrary request sequence for the CNN problem and let $r' = r(x', y')$ be the last request in σ' .

Lemma 14 *Any $(x, y) \in \mathbb{M}$ is dominated w.r.t. σ' by (x', y) or by (x, y') . (Used in proof of Lemma 17 and 19.)*

Proof: Any point is dominated by a point of the last request. Therefore, (x, y) is dominated by (x', \hat{y}) or by (\hat{x}, y') for some $\hat{y} \in \mathbb{Y}$ or $\hat{x} \in \mathbb{X}$. In general, if s is dominated by t , then s is dominated by any point on the shortest path between s and t . Now, note that (x', y) is on the shortest path between (x, y) and (x', \hat{y}) , and that (x, y') is on the shortest path between (x, y) and (\hat{x}, y') . \square

Let σ' be followed by request $r'' = r(x'', y'')$ and denote the extended sequence by $\sigma'' = \sigma', r''$. To simplify notation we denote $d_{\mathbb{X}}(x', x'') = |x' - x''|$ by ∂x and do the same for y . See Figure 3. From now on we assume without loss of generality that

$$\partial x \geq \partial y.$$

Remember the definition of extended cost. From Equation (3) we know that

$$\nabla_{r''}(W_{\sigma'}) = \max_{s \in \mathbb{M}} Sl_{\sigma'}(s; r'').$$

Since this is the only extended cost that we will consider in this proof we denote it simply by ∇ . Further, let $\xi \in \mathbb{M}$ be a point where the maximum is attained,

i.e.,

$$\nabla = \nabla_{r''}(W_{\sigma'}) = Sl_{\sigma'}(\xi; r''). \quad (8)$$

Point ξ will be used in Lemma 18 and 19.

Lemma 15 $\nabla \leq (1 + \lambda)\partial x$. (Used in proof of Lemma 19.)

Proof: Any point $s \in \mathbb{M}$ is dominated by a point in r' w.r.t. σ' . Hence, by Lemma 6(b), we may restrict to r' , i.e., $\nabla = \max_{s \in \mathbb{M}} Sl_{\sigma'}(s; r'') = \max_{s \in r'} Sl_{\sigma'}(s; r'')$. For any point s in r' there is a point in r'' at distance at most ∂x implying $Sl_{\sigma'}(s; r'') \leq (1 + \lambda)\partial x$ for any point s in r' . \square

3.4 The potential applied to CNN

In this section we apply our potential function to the CNN problem. Lemmas 18, 19 and 20 state how $\min \mathcal{F}$ and $\min \mathcal{G}$ increase when new request r'' arrives, i.e., when going from σ' to σ'' . Then, Lemma 21 combines these results and gives a lower bound on the increase of the potential function in terms of the extended cost. The proof of Theorem 1 is then straightforward.

The following lemma is given without proof as it is easy to check by looking at the definitions.

Lemma 16 Consider request sequence σ' as fixed and consider the next request $r'' = r(x'', y'')$ as a variable. Then $\min \mathcal{G}_{\sigma', r''}$ and ∇ are Lipschitz continuous in both x'' and y'' . (Used in proof of Lemma 19.)

Lemma 16 is true as well for \mathcal{F} and \mathcal{H} but we do not need that. We shall use the following easy property several times.

$$W_{\sigma'}(s) = W_{\sigma''}(s), \text{ for any } s \in r''. \quad (9)$$

Any $s \in \mathbb{M}$ is dominated w.r.t. σ' by a point in r' and Lemma 14 gives two candidate points. The next lemma reduces this to one candidate in certain cases.

Lemma 17 Assume that $\mathcal{F}_{\sigma''}$ or $\mathcal{G}_{\sigma''}$ is minimized in (s_1, s_2, s_3) . Then, the following is true for any $i \in \{1, 2, 3\}$.

1. If $s_i = (x'', y)$ for some $y \neq y'$, then (x', y) dominates s_i w.r.t. σ' .
2. If $s_i = (x, y'')$ for some $x \neq x'$, then (x, y') dominates s_i w.r.t. σ' .

(Used in proof of Lemma 18 and 19.)

Proof: We start with the first. By Lemma 14, point $s_i = (x'', y)$ is dominated w.r.t. σ' by (x'', y') or by (x', y) . Suppose the first is true. Then, using (9), s_i is dominated by this point w.r.t. σ'' as well. In that case Lemma 7 implies that $\mathcal{F}_{\sigma''}$ and $\mathcal{G}_{\sigma''}$ are strictly reduced by replacing s_i by (x'', y') . This contradicts the assumption of minimality. Thus, s_i is dominated by (x', y) w.r.t. σ' . The second case is similar. \square

We now give some easy equations that we use in the following lemmas. Equation (9) implies that if $s_1, s_2, s_3 \in r''$ then,

$$\mathcal{H}_{\sigma'}(s_1, s_2) = \mathcal{H}_{\sigma''}(s_1, s_2) \text{ and } \mathcal{F}_{\sigma'}(s_1, s_2, s_3) = \mathcal{F}_{\sigma''}(s_1, s_2, s_3).$$

(This is not true for \mathcal{G} .) Consequently, Lemma 8 implies

$$\min \mathcal{H}_{\sigma'} \leq \min \mathcal{H}_{\sigma''} \text{ and } \min \mathcal{F}_{\sigma'} \leq \min \mathcal{F}_{\sigma''}.$$

From now, we let

$$\alpha \leq \frac{1 - \lambda}{12 + 4\lambda}. \quad (10)$$

We use this bound for Lemma 18 and Lemma 19, although for Lemma 18 we could do with a weaker bound.

Lemma 18 *Let $\mathcal{F}_{\sigma''}$ be minimized in (s_1, s_2, s_3) . There are constants $c_1, c_2 > 0$ (depending on λ) such that,*

(Case A) if the cardinality of $\{s_1, s_2, s_3\}$ is 1 or 2 then

$$\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} \geq c_1 \nabla, \text{ and}$$

(Case B) if the cardinality of $\{s_1, s_2, s_3\}$ is 3 then,

$$\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} \geq c_2 \partial y.$$

(Used in proof of Lemma 21.)

Proof: Lemma 8 tells us that $s_1, s_2, s_3 \in r''$. We use this in both cases.

Case A: By Lemma 10 there are points $u_1, u_2 \in \{s_1, s_2, s_3\}$ such that $\mathcal{H}_{\sigma''}(u_1, u_2) \leq \mathcal{F}_{\sigma''}(s_1, s_3, s_3)$. Remember the definition of ξ in (8).

$$\begin{aligned} \min \mathcal{F}_{\sigma''} &= \mathcal{F}_{\sigma''}(s_1, s_2, s_3) \\ &\geq \mathcal{H}_{\sigma''}(u_1, u_2) \\ &= \mathcal{H}_{\sigma'}(u_1, u_2) \\ &= \mathcal{F}_{\sigma'}(u_1, u_2, \xi) + \alpha Sl_{\sigma'}(\xi; \{u_1, u_2\}) \\ &\geq \min \mathcal{F}_{\sigma'} + \alpha Sl_{\sigma'}(\xi; \{u_1, u_2\}) \\ &\geq \min \mathcal{F}_{\sigma'} + \alpha Sl_{\sigma'}(\xi; r'') \\ &= \min \mathcal{F}_{\sigma'} + \alpha \nabla. \end{aligned}$$

Case B: Since the points are different, at least one of the three points differs from both (x'', y') and (x', y'') . By Lemma 17, this point is dominated w.r.t. σ' by a point at distance ∂x or ∂y . Now we use Lemma 7 with $\delta = \partial y$. Note that by our bound on α all righthand sides in Lemma 7 are at least $\alpha(1 - \lambda)\partial y$.

$$\begin{aligned} \min \mathcal{F}_{\sigma''} &= \mathcal{F}_{\sigma''}(s_1, s_2, s_3) \\ &= \mathcal{F}_{\sigma'}(s_1, s_2, s_3) \\ &\geq \min \mathcal{F}_{\sigma'} + \alpha(1 - \lambda)\partial y. \end{aligned}$$

□

Lemma 19 $\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} \geq c_3 \nabla - c_4 \partial y$, for some constants $c_3, c_4 > 0$ depending on λ . (Used in proof of Lemma 21.)

Proof: By Lemma 16 it is enough if we assume that $y' = y''$ and then prove that $\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} \geq c_3 \nabla$. So we assume $y' = y''$ and it is convenient to denote both by y^* .

Let $\mathcal{G}_{\sigma''}$ be minimized for (s_1, s_2, s_3) and let $s_i = (x_i, y_i)$, for $i \in \{1, 2, 3\}$. We make the following partition of possible cases.

$$\begin{aligned} \text{Case 1:} \quad & y_1 = y^* \quad \text{and} \quad y_2 = y^* \quad \text{and} \quad y_3 = y^*, \\ \text{Case 2:} \quad & y_1 = y^* \quad \text{and} \quad y_2 = y^* \quad \text{and} \quad y_3 \neq y^*, \\ \text{Case 3:} \quad & y_1 \neq y^* \quad \text{or} \quad y_2 \neq y^*. \end{aligned}$$

By Lemma 8, we have $s_1, s_2, s_3 \in r''$. We shall use this property several times here. For example, if $y_i \neq y^*$ then $x_i = x''$.

Case 1: We apply Lemma 11: $\mathcal{H}_{\sigma''}(u_1, u_2) \leq \mathcal{G}_{\sigma''}(s_1, s_2, s_3)$ for some $u_1, u_2 \in \{s_1, s_2, s_3\}$.

$$\begin{aligned} \min \mathcal{G}_{\sigma''} &= \mathcal{G}_{\sigma''}(s_1, s_2, s_3) \\ &\geq \mathcal{H}_{\sigma''}(u_1, u_2) \\ &= \mathcal{H}_{\sigma'}(u_1, u_2) \\ &= \mathcal{G}_{\sigma'}(u_1, u_2, \xi) + \alpha Sl_{\sigma'}(\xi; \text{BOX}(u_1, u_2)) \\ &\geq \min \mathcal{G}_{\sigma'} + \alpha Sl_{\sigma'}(\xi; \text{BOX}(u_1, u_2)) \\ &\geq \min \mathcal{G}_{\sigma'} + \alpha Sl_{\sigma'}(\xi; r'') \\ &= \min \mathcal{G}_{\sigma'} + \alpha \nabla. \end{aligned}$$

The last inequality follows from $\text{BOX}(u_1, u_2) \subset r''$ and Lemma 2.

Case 2: Since $\text{BOX}(s_1, s_2) \subset r''$ and $s_1, s_2, s_3 \in r''$ we have

$$\mathcal{G}_{\sigma''}(s_1, s_2, s_3) = \mathcal{G}_{\sigma'}(s_1, s_2, s_3).$$

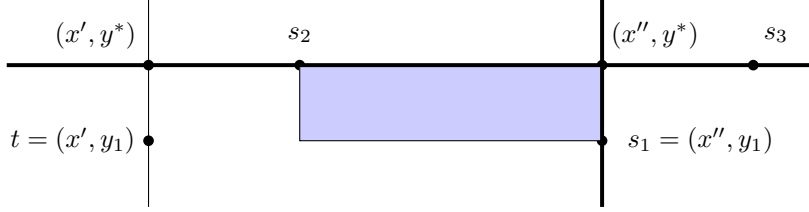


Figure 4: Case 3 of Lemma 19: $y_1 \neq y^*$. The shaded area is $\text{Box}(s_1, s_2)$.

By Lemma 17, point $s_3 = (x'', y_3)$ is dominated by point $t = (x', y_3)$ with respect to σ' . Now we apply Lemma 7-d.

$$\begin{aligned}
\min \mathcal{G}_{\sigma''} &= \mathcal{G}_{\sigma''}(s_1, s_2, s_3) \\
&= \mathcal{G}_{\sigma'}(s_1, s_2, s_3) \\
&\geq \mathcal{G}_{\sigma'}(s_1, s_2, t) + \alpha(1 - \lambda)\partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \alpha(1 - \lambda)\partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \alpha(1 - \lambda)\nabla/(1 + \lambda).
\end{aligned}$$

The last inequality is given by Lemma 15.

Case 3: See Figure 4. Unlike the previous two cases, we may now have $\text{Box}(s_1, s_2) \not\subseteq r''$ which makes the proof slightly more complicated. W.l.o.g we may assume that $y_1 \neq y^*$. This implies $x_1 = x''$ and point $s_1 = (x'', y_1)$ is dominated by point $t = (x', y_1)$ with respect to σ' . We now apply Lemma 7(f).

$$\begin{aligned}
\mathcal{G}_{\sigma'}(s_1, s_2, s_3) &\geq \mathcal{G}_{\sigma'}(t, s_2, s_3) + \left(\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda)\right) \partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \left(\frac{1}{2}(1 - \lambda) - \alpha(1 + \lambda)\right) \partial x
\end{aligned} \tag{11}$$

It remains to bound $\min \mathcal{G}_{\sigma''} - \mathcal{G}_{\sigma'}(s_1, s_2, s_3)$. We have ²

$$\begin{aligned}
&\min \mathcal{G}_{\sigma''} - \mathcal{G}_{\sigma'}(s_1, s_2, s_3) \\
&\mathcal{G}_{\sigma''}(s_1, s_2, s_3) - \mathcal{G}_{\sigma'}(s_1, s_2, s_3) \\
&= \alpha Sl_{\sigma'}(s_3; \text{Box}(s_1, s_2)) - \alpha Sl_{\sigma''}(s_3; \text{Box}(s_1, s_2)) \\
&\geq -2\alpha\partial x.
\end{aligned} \tag{12}$$

The last inequality follows from $W_{\sigma'}(s_3) = W_{\sigma''}(s_3)$ and from $W_{\sigma''}(s) - W_{\sigma'}(s) \leq 2\partial x$ for any point $s \in \mathbb{M}$ (and $s \in \text{Box}(s_1, s_2)$ in particular). Below we use,

²A more careful analysis gives a bound $-(1 + \lambda)\alpha\partial x$ in stead of $-2\alpha\partial x$.

subsequently, (12), (11), (10) and Lemma 15.

$$\begin{aligned}
& \min \mathcal{G}_{\sigma''} \\
&= \mathcal{G}_{\sigma''}(s_1, s_2, s_3) \\
&\geq \mathcal{G}_{\sigma'}(s_1, s_2, s_3) - 2\alpha \partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \left(\frac{1}{2}(1-\lambda) - \alpha(1+\lambda)\right) \partial x - 2\alpha \partial x \\
&= \min \mathcal{G}_{\sigma'} + \left(\frac{1}{2}(1-\lambda) - \alpha(3+\lambda)\right) \partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \left(\frac{1}{2}(1-\lambda) - \frac{1}{4}(1-\lambda)\right) \partial x \\
&= \min \mathcal{G}_{\sigma'} + \frac{1}{4}(1-\lambda) \partial x \\
&\geq \min \mathcal{G}_{\sigma'} + \frac{1}{4}(1-\lambda) \nabla / (1+\lambda).
\end{aligned}$$

This completes the proof of the last case. \square

In Lemma 21 we combine Lemmas 18 and 19 and distinguish the same two cases A and B as we did in Lemma 18. Lemma 19 will be used only for Case B, although it holds in general. For Case A we need the following different bound.

Lemma 20 *Let $\mathcal{F}_{\sigma''}$ be minimized in (s_1, s_2, s_3) . If the cardinality of (s_1, s_2, s_3) is 1 or 2 then $\min \mathcal{G}_{\sigma''} \geq \min \mathcal{G}_{\sigma'}$. (Used in proof of Lemma 21.)*

Proof: By Lemma 10 there are points $u_1, u_2 \in \{s_1, s_2, s_3\}$ such that $\min \mathcal{H}_{\sigma''} \leq \mathcal{H}_{\sigma''}(u_1, u_2) \leq \mathcal{F}_{\sigma''}(s_1, s_2, s_3) = \min \mathcal{F}_{\sigma''}$. In Lemma 9, the inequalities are the other way around. Hence,

$$\min \mathcal{F}_{\sigma''} = \min \mathcal{G}_{\sigma''} = \min \mathcal{H}_{\sigma''}.$$

We conclude that

$$\min \mathcal{G}_{\sigma''} = \min \mathcal{H}_{\sigma''} \geq \min \mathcal{H}_{\sigma'} \geq \min \mathcal{G}_{\sigma'}.$$

\square

Lemma 21 $\Phi_{\sigma''} - \Phi_{\sigma'} \geq c_5 \nabla$ for some constant $c_5 > 0$, depending on λ . (Used in proof of Theorem 1.)

Proof:

$$\Phi_{\sigma''} - \Phi_{\sigma'} = (1-\gamma)(\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'}) + \gamma(\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'})$$

Let $\mathcal{F}_{\sigma''}$ be minimized in (s_1, s_2, s_3) . We distinguish between the same two cases as in Lemma 18:

Case A: The cardinality of $\{s_1, s_2, s_3\}$ is 1 or 2.

Case B: The cardinality of $\{s_1, s_2, s_3\}$ is 3.

Case A: By Lemma 18 $\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} \geq c_1 \nabla$, for some constant $c_1 > 0$ and by Lemma 20 $\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} \geq 0$. Hence,

$$\Phi_{\sigma''} - \Phi_{\sigma'} \geq (1 - \gamma)c_1 \nabla.$$

Case B: By Lemma 18 and Lemma 19

$$\begin{aligned} \min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'} &\geq c_2 \partial y, \\ \min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} &\geq c_3 \nabla - c_4 \partial y, \end{aligned}$$

for some constants $c_2, c_3, c_4 > 0$. Hence,

$$\begin{aligned} \Phi_{\sigma''} - \Phi_{\sigma'} &= (1 - \gamma)(\min \mathcal{F}_{\sigma''} - \min \mathcal{F}_{\sigma'}) + \gamma(\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'}) \\ &\geq (1 - \gamma)c_2 \partial y + \gamma(c_3 \nabla - c_4 \partial y) \\ &= \gamma c_3 \nabla + ((1 - \gamma)c_2 - \gamma c_4) \partial y. \end{aligned}$$

By choosing γ small enough the constant before ∂y will be positive. We choose $(1 - \gamma) = \gamma c_4 / c_2$, i.e. $\gamma = c_2 / (c_2 + c_4)$. Hence,

$$\Phi_{\sigma''} - \Phi_{\sigma'} \geq \gamma c_3 \nabla. \tag{13}$$

Combining Case A and Case B we obtain

$$\Phi_{\sigma''} - \Phi_{\sigma'} \geq \min \{(1 - \gamma)c_1, \gamma c_3\} \nabla = c_5 \nabla,$$

where

$$c_5 = \min \{(1 - \gamma)c_1, \gamma c_3\} = \min \left\{ \frac{c_1 c_4}{c_2 + c_4}, \frac{c_2 c_3}{c_2 + c_4} \right\}.$$

□

Proof of Theorem 1: Let ρ be any request sequence. Using Lemma 21 and taking the sum over all requests we get

$$\Phi_\rho - \Phi_\epsilon \geq c_5 \nabla_\rho.$$

Lemma 12 states that $\Phi_\epsilon = 0$, and Lemma 13 states that $\Phi_\rho \leq \text{OPT}_\rho$. Hence,

$$\nabla_\rho \leq \frac{1}{c_5} \text{OPT}_\rho.$$

By Lemma 1, the competitive ratio is at most $(1/c_5 - 1)/\lambda$. □

4 General metric spaces

In this section we extend Theorem 1 to arbitrary symmetric metric spaces.

Theorem 2 *The work function algorithm WFA_λ is constant competitive for the generalized 2-server problem for any constant λ with $0 < \lambda < 1$.*

The work function algorithm WFA_λ is well-defined for the generalized 2-server problem on any metric space and request sequence σ since the only interesting points are (x, y) where x and y are points in \mathbb{X} and \mathbb{Y} that were requested so far in σ . These are also the only interesting points for our potential function. However, in the analysis we do use that the potential function is a Lipschitz continuous function of the given request (Lemma 16) and this is not true in general in a discrete metric space. We need to extend the metric space into a convex space $\overline{\mathbb{M}} \supseteq \mathbb{M}$ where any two points are joined by a continuous path, i.e., for any pair $u_1, u_2 \in \overline{\mathbb{M}}$ and $\zeta \in [0, 1]$ there is a point $u_3 \in \overline{\mathbb{M}}$ such that $d(u_1, u_3) = \zeta d(u_1, u_2)$ and $d(u_2, u_3) = (1 - \zeta)d(u_1, u_2)$. This can easily be done and is a common assumption for online routing problems. See for example [11] for a discussion on this. We avoid using the notation $\overline{\mathbb{M}}$ and simply assume that \mathbb{M} has this property. Note that this is done only for the analysis. The request sequence and the work function algorithm will only use points of the original metric space.

On one hand, the generalization of the proof is easy since all lemmas stay exactly the same, apart from some constants. But to achieve that we need to make the potential function even more complicated than it is. Then again, the good thing is that the only proof that really changes is that of Lemma 11.

4.1 Adjusting the potential

The first point in the CNN proof where we used the restriction to \mathbb{R}^2 is in the potential function: The set $\text{BOX}(s_1, s_2)$ is defined only for $s_1, s_2 \in \mathbb{R}^2$. Why did we need this definition of BOX? It was defined especially for Lemma 11: If the points (s_1, s_2, s_3) have the same x - or y -coordinate, then one of them is redundant. We applied this in Lemma 19 (Case 1) where we replaced the redundant point by ξ . The proof of Lemma 11 doesn't apply here because of Equality (7). In general, the equality is an inequality:

$$Sl_\sigma(s_3; s_1) \leq Sl_\sigma(s_3; s_2) + Sl_\sigma(s_2; s_1).$$

Unfortunately, we need \geq here for the proof to hold. Looking ahead at Equation (16) one sees that we somehow managed to do this. The trick is simple. We make two changes to the potential function: We add the constraint that s_3

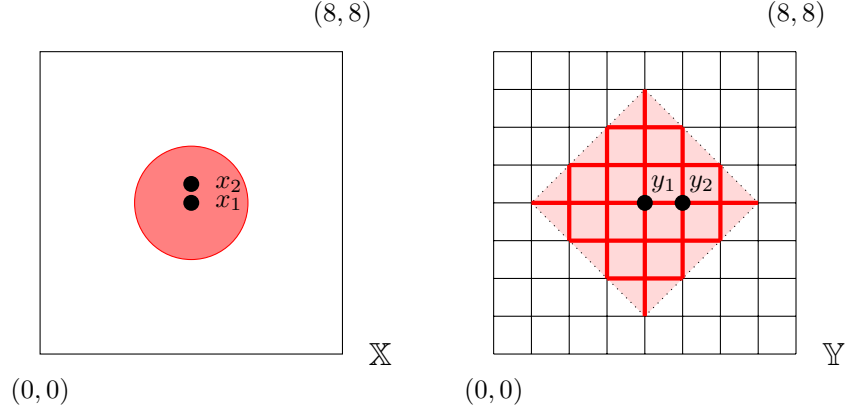


Figure 5: Example of SPHERES. Here, \mathbb{X} is the Euclidean plane and \mathbb{Y} is the unit grid with the L_1 norm. Further, $s_1 = (x_1, y_1) = ((4, 4), (4, 4))$, and $s_2 = (x_2, y_2) = ((4, 4\frac{1}{2}), (5, 4))$. Hence, $d^{\mathbb{X}}(x_1, x_2) = 0.5$ and $d^{\mathbb{Y}}(y_1, y_2) = 1$. Constant $\eta = 3$ and $\text{SPHERES}(s_1, s_2)$ is the Cartesian product of the shaded areas. The unit grid \mathbb{Y} is extended with the gridlines, i.e., we assume grid points are connected by continuous paths. See the remark in the beginning of Section 4.1

should be relatively far from s_1 and s_2 and we take two different measures for slack, i.e., the slack for s_3 is taken less steep than the slack for s_2 . We make this precise below.

The following definition takes the place of BOX. Let $\eta \gg 1$.

$$\text{SPHERES}(s_1, s_2) = \{ (x, y) \in \mathbb{X} \times \mathbb{Y} \mid d^{\mathbb{X}}(x, x_1) \leq \eta \cdot d^{\mathbb{X}}(x_1, x_2) \\ \text{and } d^{\mathbb{Y}}(y, y_1) \leq \eta \cdot d^{\mathbb{Y}}(y_1, y_2) \}.$$

Note that SPHERES is in fact the Cartesian product of a sphere around x_1 and a sphere around y_1 . Instead of (x_1, y_1) we could also take (x_2, y_2) or somehow a point in between. This makes no real difference if η is large. (One could think of $\text{BOX}(s_1, s_2)$ as the Cartesian product of a 1-dimensional sphere of diameter $|x_2 - x_1|$ around point $(x_1 + x_2)/2$ and a 1-dimensional sphere of diameter $|y_2 - y_1|$ around point $(y_1 + y_2)/2$.)

Another change that we make in the potential function is adjusting the constants. The whole proof for Theorem 1 is still valid (up to a constant) if we replace the λ 's that appear in the potential function by some other constant μ for which $\lambda \leq \mu < 1$, while keeping WFA_λ the same. There is no need to verify this claim since we do not use it explicitly but it is the reason for finetuning the potential function as we will do here. Let us fix such a μ with $\lambda < \mu < 1$ and

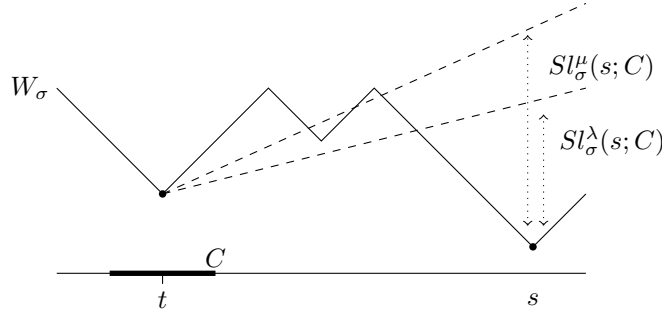


Figure 6: Two kinds of slack: one with parameter λ and one with parameter $\mu > \lambda$.

define for $s \in \mathbb{M}$ and $C \subseteq \mathbb{M}$ the slack as before but with μ in stead of λ . In addition we keep the old definition and add the parameter λ in the notation.

$$Sl_\sigma^\mu(s; C) = \min_{t \in C} \{W_\sigma(t) + \mu d(t, s)\} - W_\sigma(s)$$

$$Sl_\sigma^\lambda(s; C) = \min_{t \in C} \{W_\sigma(t) + \lambda d(t, s)\} - W_\sigma(s).$$

Next we define the new \mathcal{H}_σ , \mathcal{F}_σ and \mathcal{G}_σ . For simplicity we keep these names as before although they are now slightly different functions.

$$\mathcal{H}_\sigma(s_1, s_2) = W_\sigma(s_1) - \frac{1}{2} Sl_\sigma^\mu(s_2; s_1)$$

$$\mathcal{F}_\sigma(s_1, s_2, s_3) = \mathcal{H}_\sigma(s_1, s_2) - \beta Sl_\sigma^\lambda(s_3; \{s_1, s_2\})$$

$$\mathcal{G}_\sigma(s_1, s_2, s_3) = \mathcal{H}_\sigma(s_1, s_2) - \beta Sl_\sigma^\lambda(s_3; \text{SPHERES}(s_1, s_2)).$$

Note that μ is used for the slack of s_2 while λ is used for the slack of s_3 . The potential function is:

$$\Phi_\sigma = (1 - \kappa) \min \mathcal{F}_\sigma + \kappa \min \mathcal{G}_\sigma,$$

where $0 < \kappa < 1$. To prove constant competitiveness there is no need to specify precise values of the constants. We only need to choose the constants large or small enough. The order in which we choose them and the domains are listed below. For example, given λ and the choice of μ there is a number η_0 such that any choice $\eta \geq \eta_0$ is fine. We do not compute the values η_0 , β_0 or κ_0 but it will be clear from the proof that such values exist.

$$\begin{aligned} \lambda &: \text{ given,} \\ \mu &: \lambda < \mu < 1, \\ \eta &: \eta \geq \eta_0 \gg 1 & , \text{ where } \eta_0 \text{ depends on } \lambda \text{ and } \mu, \\ \beta &: 0 < \beta \leq \beta_0 < 1/2 & , \text{ where } \beta_0 \text{ depends on } \lambda, \mu \text{ and } \eta, \\ \kappa &: 0 < \kappa < \kappa_0 < 1 & , \text{ where } \kappa_0 \text{ depends on } \lambda, \mu, \eta \text{ and } \beta. \end{aligned}$$

4.2 Adjusting the proofs of the lemmas

All lemmas stay the same apart from some constants. Nothing changes for Section 2 since it comes before the potential function and holds for any metric space. The only proof that really changes with this new potential function is that of Lemma 11. We give a new proof below. Let us go over all the lemmas of Section 3 one by one.

Lemma 7 is still true but with different constants. We do not really need to compute precise bounds. All we need to see is that by choosing β small enough the right hand sides are strictly positive for $\delta > 0$. The bounds we get are

$$\begin{aligned} (a), (d) &: \delta \cdot \beta(1 - \lambda) \\ (b), (c) &: \delta \cdot \left(\frac{1}{2}(1 - \mu) - \beta(1 + \lambda)\right) \\ (e) &: \delta \cdot \left(\frac{1}{2}(1 - \mu) - \eta\beta(1 + \lambda)\right) \\ (f) &: \delta \cdot \left(\frac{1}{2}(1 - \mu) - (\eta + 1)\beta(1 + \lambda)\right) \end{aligned}$$

The proof for (a)-(d) is the same. In (e) there is an additional factor η because a move of s_2 over some distance may cause the border of $\text{SPHERES}(s_1, s_2)$ to move by η times this distance. For a move of s_1 this factor is $\eta + 1$ since SPHERES is defined around s_1 . Now we choose β small enough to let all the right hand sides be strictly positive for $\delta > 0$.

Lemmas 8, 9 and 10 hold by similar arguments. In Lemma 9 we simply replace BOX by SPHERES and in Lemma 10 we only need to update the definition of \mathcal{F} .

New proof of Lemma 11: Let s_1, s_2, s_3 have the same y -coördinate. We may assume that

$$Sl_\sigma^\lambda(s_3; \text{SPHERES}(s_1, s_2)) > 0, \quad (14)$$

since otherwise $\mathcal{H}_\sigma(s_1, s_2) \leq \mathcal{G}_\sigma(s_1, s_2, s_3)$ and we are done. By this assumption we have $s_3 \notin \text{SPHERES}(s_1, s_2)$. Hence, $d(s_1, s_3) > \eta d(s_1, s_2)$ (using $d^\mathbb{X}(s_i, s_j) = d(s_i, s_j)$ for $i, j \in \{1, 2, 3\}$). Then

$$\begin{aligned} Sl_\sigma^\mu(s_3; s_1) &= W_\sigma(s_1) + \mu d(s_1, s_3) - W_\sigma(s_3) \\ &= W_\sigma(s_1) + \lambda d(s_1, s_3) - W_\sigma(s_3) + (\mu - \lambda)d(s_1, s_3) \\ &= Sl_\sigma^\lambda(s_3; s_1) + (\mu - \lambda)d(s_1, s_3) \\ &> Sl_\sigma^\lambda(s_3; s_1) + \eta(\mu - \lambda)d(s_1, s_2). \end{aligned} \quad (15)$$

By choosing η large enough (given the values of λ and μ) we guarantee that $\eta(\mu - \lambda) \geq 1 + \mu$. If we also use that $(1 + \mu)d(s_1, s_2) \geq Sl_\sigma^\mu(s_2; s_1)$ (follows directly from (14)) then (15) becomes

$$Sl_\sigma^\mu(s_3; s_1) > Sl_\sigma^\lambda(s_3; s_1) + Sl_\sigma^\mu(s_2; s_1). \quad (16)$$

The remainder of the proof is similar to the original proof. For the first two inequalities below we use, respectively, (16) and Lemma 2. For the last inequality we use (14) and $\beta < 1/2$.

$$\begin{aligned}
\mathcal{H}_\sigma(s_1, s_3) &= W_\sigma(s_1) - \frac{1}{2}Sl_\sigma^\mu(s_3; s_1) \\
&< W_\sigma(s_1) - \frac{1}{2}Sl_\sigma^\mu(s_2; s_1) - \frac{1}{2}Sl_\sigma^\lambda(s_3; s_1) \\
&\leq W_\sigma(s_1) - \frac{1}{2}Sl_\sigma^\mu(s_2; s_1) - \frac{1}{2}Sl_\sigma^\lambda(s_3; \text{SPHERES}(s_1, s_2)) \\
&< W_\sigma(s_1) - \frac{1}{2}Sl_\sigma^\mu(s_2; s_1) - \beta Sl_\sigma^\lambda(s_3; \text{SPHERES}(s_1, s_2)) \\
&= \mathcal{G}_\sigma(s_1, s_2, s_3).
\end{aligned}$$

□

Lemmas 12, and 13 remain the same and follow by exactly the same proofs. Only now α becomes β and γ becomes κ . Lemmas 14 and 15 do not depend on the potential function. These lemmas and proofs stay exactly the same. Lemma 16 can again easily be verified from the definitions. Lemma 17 does depend on \mathcal{F} and \mathcal{G} but also here the lemma and proof remain exactly the same. In Case A of the proof of Lemma 18 the only change is that α becomes β . In Case B, the last inequality is different since the inequalities of Lemma 7 are different. (The new values were given earlier in this section.) All we need from Lemma 7 is that by choosing β small enough the righthand sides are strictly positive.

Also Lemma 19 remains basically the same. Of course, α becomes β and BOX becomes SPHERES. The new function \mathcal{G} is still Lipschitz continuous. Hence we may assume $y' = y''$. We consider the same three cases and the proof for the first and second case remain the same. For Case 3 we need to use the new bounds of Lemma 7. Then, Equation (11) becomes

$$\mathcal{G}_{\sigma'}(s_1, s_2, s_3) \geq \min \mathcal{G}_{\sigma'} + \left(\frac{1}{2}(1 - \mu) - (\eta + 1)\beta(1 + \lambda) \right) \partial x.$$

Combining this with Equation (12) as we did we get

$$\min \mathcal{G}_{\sigma''} - \min \mathcal{G}_{\sigma'} \geq \left(\frac{1}{2}(1 - \mu) - (\eta + 1)\beta(1 + \lambda) - 2\beta \right) \partial x.$$

By choosing β small enough the righthand side is at least $c_3 \nabla$ for some constant c_3 (depending on λ, μ and η). Finally, the proof of Lemma 20 remains the same and for Lemma 21 only the γ becomes κ .

5 Remarks

The obvious open problem is to reduce the gap between the upper and lower bound of the generalized 2-server problem or even the CNN problem. A better understanding of the work function for the CNN problem is essential. Our proof uses very little of the structure of the work function (Section 3.3). The power comes from a sophisticated potential function, one that has all the needed properties, and not because it matches the CNN problem so well but because we simply constructed the function to let it have all the properties we want. But this is also a good thing. Our potential function has some ingredients that are interesting for the analysis of work function based algorithms in general. For example, using some convex set like BOX and SPHERES and taking the slack to that set and, for example, using slack functions with different parameters, e.g. λ and μ .

5.1 Higher dimensions

The k -CNN problem seems much more complicated for dimensions $k \geq 3$ than the 2-CNN problem. It is unclear if a constant competitive ratio is possible at all for $k \geq 3$. The question is interesting for its relation to sum problems, discussed in the introduction. In any case, the ratio will be at least $k^{\Omega(k)}$ [16]. We give an outline for a possible extension of our proof to higher dimensions. The proof we gave here follows this outline for $k = 2$.

The potential function is

$$\Phi_\sigma = \sum_{i=1}^k \gamma^{(i)} \cdot \min_{s_1, \dots, s_{k+1} \in \mathbb{M}} \mathcal{F}_\sigma^{(i)}(s_1, s_2, \dots, s_{k+1}).$$

The values $\gamma^{(i)}$ are exponentially decreasing constants depending on λ . We consider the increase of the potential function over two subsequent requests $(x'_1, x'_2, \dots, x'_k)$ and $(x''_1, x''_2, \dots, x''_k)$. The first function, $\min \mathcal{F}^{(1)}$, has the property that the increase is either $O(\nabla)$ or it is $O(\min_i \partial x_i)$. In the first case we are done if in this case $\min \mathcal{F}^{(i)}$ is non-decreasing for all $i \geq 2$. Then, for the second case we may assume from now that $\partial x_{j_1} = 0$ for some j_1 with $\partial x_{j_1} = \min_i \partial x_i$. (We require that the functions $\min \mathcal{F}^{(i)}$ are Lipschitz continuous in r'' .) Hence, we get rid of one dimension. Again there are two cases: either the increase for $\min \mathcal{F}^{(2)}$ is $O(\nabla)$ or it is $O(2\text{nd smallest value of } \partial x_i)$. In the first case we are done if in this case $\min \mathcal{F}^{(i)}$ is non-decreasing for all $i \geq 3$. In the second case we may assume from now that $\partial x_{j_2} = 0$ for another value j_2 . Again we get rid of one dimension. This goes on until finally for $\min \mathcal{F}^{(k)}$ we may assume that

the two subsequent requests differ in only one coördinate and we can show that $\min \mathcal{F}^{(k)}$ is $O(\nabla)$.

5.2 Open problems

There are some very intriguing open problems in online optimization. Examples are the k -server conjecture (deterministic and randomized) and the dynamic search tree conjecture [30]. Dynamic search trees (without insertion or deletions) are in the class of metrical service systems and it is a general believe that dynamic search trees are constant competitive. In this paper we showed some strong techniques for proving constant competitiveness of metric service systems which might be applicable to dynamic search trees as well.

Below, we list some interesting open problems related to the CNN problem.

1. Give a constant competitive work function based algorithm for dynamic search trees (without insertions or deletions). Although the algorithm would be inefficient it would clearly be a big step towards more efficient constant competitive algorithm like splaytrees.
2. Prove or disprove that the k -CNN or weighted k -server problem has an $f(k)$ -competitive algorithm for some function $f(k)$. Same for the randomized problem.
3. What is the right competitive ratio of the k -point request problem? Burley [8] gives an upper bound of $O(k2^k)$ while the best known lower bound is $O(2^k)$. Just as Burley we conjecture that $O(2^k)$ is possible.
4. What is the competitive ratio of the continues CNN problem? A lower bound of 3 and upper bound of 6.46 is given in [1].
5. Give other examples of natural metrical service systems that have a constant competitive ratio. For example, Friedman and Linial [17] give a competitive algorithm if the requests are convex subset of \mathbb{R}^2 . They conjecture that the same applies to \mathbb{R}^d for any fixed d and show that it is enough to prove this for affine half planes.
6. The k -server problem has some simple special cases for which $2k - 1$ is still the best known ratio, for example the 3-server problem and the k -server problem on a cycle: Find an algorithm with a smaller ratio. For ratio for trees is k but it is unknown if the work function algorithm achieves this ratio. See [21] for more background on this.
7. What is the competitive ratio of the weighted k -point request problem, discussed in [15]?

8. Extend the theory of sum problems. For example by analyzing the sum problem of another elementary metrical task system.
9. Prove (or disprove) that the generalized work function algorithm WFA_λ is $O(\log n)$ -competitive for the online matching problem on a line.[22]

References

- [1] J. Augustine and N. Gravin. On the continuous CNN problem. In *Proc. 21th International Symp. on Algorithms and Computation*, volume 6507 of *Lecture Notes in Computer Science*, pages 254–265. Springer, 2010.
- [2] G. Ausiello, V. Bonifaci, and L. Laura. Lazy on-line algorithms for metrical service systems. Technical Report DIS 07-04, Department of Computer and System Sciences, Università di Roma La Sapienza, Roma, Italy, 2004.
- [3] G. Ausiello, V. Bonifaci, and L. Laura. On explorers, chasers and cameramen. In *Proc. of 3rd int. conf. on fun with algorithms (FUN2004)*, 2004.
- [4] Y. Bartal and E. Koutsoupias. On the competitive ratio of the work function algorithm for the k -server problem. *Theoretical Computer Science*, 324:337–345, 2004.
- [5] A. Blum and C. Burch. On-line learning and the metrical task system problem. *Machine Learning*, 39:35–58, 2000.
- [6] A. Borodin, N. Linial, and M. Saks. An optimal online algorithm for metrical task system. *Journal of the ACM*, 39:745–763, 1992.
- [7] A. Borodin and E.-Y. R. *Online computation and competitive analysis*. Cambridge University Press, 1998.
- [8] W. Burley. Traversing layered graphs using the work function algorithm. *Journal of Algorithms*, 20:479–511, 1996.
- [9] M. Chrobak. Column on online algorithms. *SIGACT News*, 34:68–77, 2003.
- [10] M. Chrobak and L. Larmore. An optimal online algorithm for k servers on trees. *SIAM Journal on Computing*, 20:144–148, 1991.
- [11] M. Chrobak and L. Larmore. The server problem and on-line games. In *On-line algorithms: Proceedings of a DIMACS Workshop*, volume 7 of *Dimacs Series in Discrete Mathematics and Theoretical Computer Science*, pages 11–64, 1992.

- [12] M. Chrobak and L. Larmore. Metrical service systems: Deterministic strategies. Technical Report UCR-CS-93-1, Department of Computer Science, University of California at Riverside, 1993.
- [13] M. Chrobak and L. Larmore. Metrical task systems, the server problem and the work function algorithm. In A. Fiat and G. Woeginger, editors, *Online algorithms: the state of the art*, volume 1442 of *Lecture Notes in Computer Science*, chapter 4, pages 74–96. Springer, 1998.
- [14] M. Chrobak and J. Sgall. The weighted 2-server problem. In *Proc. 17th Symp. on Theoretical Aspects of Computer Science*, volume 1770 of *Lecture Notes in Computer Science*, pages 593–604. Springer, 2000.
- [15] M. Chrobak and J. Sgall. The weighted 2-server problem. *Theoretical Computer Science*, 324:289–312, 2004.
- [16] A. Fiat and M. Ricklin. Competitive algorithms for the weighted server problem. *Theoretical Computer Science*, 130:85–99, 1994.
- [17] J. Friedman and N. Linial. On convex body chasing. *Discrete Computational Geometry*, 9:293–321, 1993.
- [18] K. Iwama and K. Yonezawa. Axis-bound CNN problem. Technical report, National Institute of Informatics Japan, 2001.
- [19] K. Iwama and K. Yonezawa. Axis-bound CNN problem. *IEICE TRANS*, E85-A:1–8, 2002.
- [20] K. Iwama and K. Yonezawa. The orthogonal CNN problem. *Information Processing Letters*, 94:115–120, 2004.
- [21] E. Koutsoupias. The k -server problem. *Computer Science Review*, 3:105–118, 2009.
- [22] E. Koutsoupias and A. Nanavati. The online matching problem on a line. In *Proc. 1st Workshop on Approximation and Online Algorithms*, pages 197–191, 2003.
- [23] E. Koutsoupias and C. Papadimitriou. On the k -server conjecture. *Journal of the ACM*, 42:971–983, 1995.
- [24] E. Koutsoupias and D. Taylor. The CNN problem and other k -server variants. In *Proc. 17th Symp. on Theoretical Aspects of Computer Science*, volume 1770 of *Lecture Notes in Computer Science*, pages 581–592. Springer, 2000.

- [25] E. Koutsoupias and D. S. Taylor. The CNN problem and other k-server variants. *Theoretical Computer Science*, 324:347–359, 2004.
- [26] M. Manasse, L. McGeoch, and D. Sleator. Competitive algorithms for on-line problems. In *Proc. 20th Symp. Theory of Computing*, pages 322–333, 1988.
- [27] M. Manasse, L. McGeoch, and D. Sleator. Competitive algorithms for server problems. *Journal of Algorithms*, 11:208–230, 1990.
- [28] R. Sitters and L. Stougie. The generalized two-server problem. *Journal of the ACM*, 53:437–458, 2006.
- [29] R. Sitters, L. Stougie, and W. de Paepe. A competitive algorithm for the general 2-server problem. In J. Baeten, J. Lenstra, J. Parrow, and G. Woeginger, editors, *Proc. 30th International Colloquium on Automata, Languages, and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 624–636. Springer, 2003.
- [30] D. Sleator and R. Tarjan. Self-adjusting binary search trees. *Journal of the ACM*, 32:652–686, 1985.
- [31] Y. F. Verhoeven. A lower bound on the competitiveness of memoryless algorithms for a generalization of the CNN problem. *Theoretical Computer Science*, 359:58–68, 2006.